

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 27, 2001, 16:52:25 ; Search time 16.97 Seconds  
(without alignments) 1456.222 Million cell updates/sec

Title: US-09-830-647-1

Perfect score: 3510  
Sequence: 1 MNSGMRIHSKHFQGGIOV.....SDNLTAFFSSPSTSTFTGF 674

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues  
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwisProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	187	5.1	1790	1 US01_YEAST	P23386 saccharomyc
2	178	5.1	3911	1 AKA9_HUMAN	O99996 h a kinase
3	165	4.7	1433	1 REST_CHICK	O42184 gallus gall
4	163.5	4.7	2230	1 GOG4_HUMAN	Q13439 homo sapien
5	159.5	4.5	1427	1 REST_HUMAN	P30622 homo sapien
6	156.5	4.5	3210	1 CENF_HUMAN	P49454 homo sapien
7	155.5	4.4	1056	1 YNN2_YEAST	P53914 saccharomyc
8	154	4.4	2663	1 CENE_HUMAN	O02224 homo sapien
9	152	4.3	1216	1 PI31_HUMAN	O90666 homo sapien
10	152	4.3	1312	1 RA50_YEAST	P13753 saccharomyc
11	152	4.3	1980	1 MY9B_NAT	Q63358 rattus norv
12	151.5	4.3	1044	1 YAF3_SCHPO	O09857 schizosacch
13	151.5	4.3	2104	1 MY53_SCHPO	O14157 schizosacch
14	149.5	4.3	1726	1 MSP1_PLAFC	P04934 plasmodium
15	149.5	4.3	1726	1 MSP1_PLAFC	P50495 plasmodium
16	149	4.2	2245	1 MY5J_DICDI	P54697 dictyostell
17	148.5	4.2	1172	1 C014_SCHPO	P41003 schizosacch
18	148.5	4.2	1805	1 HMM2_MYGE	P47460 mycoplasma
19	148.5	4.2	1875	1 MY9B_MOUSE	O94906 mus musculu
20	148	4.2	1875	1 MLP1_YEAST	O02455 saccharomyc
21	147.5	4.2	844	1 BR01_YEAST	P48582 saccharomyc
22	147.5	4.2	1978	1 MYHB_CHICK	P10587 gallus gall
23	147	4.2	1957	1 YD86_SCHPO	O10411 schizosacch
24	146.5	4.2	1216	1 PI31_NAT	P10687 rattus norv
25	146.5	4.2	1257	1 RB01_HUMAN	P29374 homo sapien
26	145.5	4.1	756	1 Y328_MYGE	O49419 mycoplasma
27	145.5	4.1	1630	1 MSP1_PLAFC	P04933 plasmodium
28	145.5	4.1	1639	1 MSP1_PLAFC	P04933 plasmodium
29	144.5	4.1	1031	1 KINH_STRPU	P35788 strongyloce
30	144.5	4.1	1225	1 SMC1_YEAST	P32308 saccharomyc
31	144.5	4.1	1226	1 YCS3_YEAST	P25357 saccharomyc
32	144.5	4.1	1290	1 XCP3_XENLA	P50532 xenopus lae
33	143.5	4.1	1616	1 P200_MYGE	O49429 mycoplasma

34	143	4.1	1087	1 AKA9_RABIT	O28628 oryctolagus
35	142.5	4.1	935	1 KINH_SYNRA	O43093 syncephalus
36	142.5	4.1	1216	1 PI31_BOVIN	P10894 bos taurus
37	142.5	4.1	1453	1 MYH3_HUMAN	O99hcl homo sapien
38	142.5	4.1	1513	1 STU1_YEAST	P38198 saccharomyc
39	142.5	4.1	1679	1 Y109_YEAST	P40457 saccharomyc
40	142	4.0	1701	1 MSP1_PLAFC	P13819 plasmodium
41	142	4.0	1701	1 MSP1_PLAFC	P08569 plasmodium
42	142	4.0	1727	1 ALM1_SCHPO	O9utk5 schizosacch
43	142	4.0	1937	1 MYHB_HUMAN	P13535 homo sapien
44	141.5	4.0	1972	1 MYHB_RABIT	P35748 oryctolagus
45	141	4.0	853	1 YCG1_YEAST	P25588 saccharomyc

## ALIGNMENTS

```

RESULT 1
US01_YEAST STANDARD: PRT: 1790 AA.
AC P23386;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE INTRACELLULAR PROTEIN TRANSPORT PROTEIN US01.
CN US01 OR INT1 OR YD1058W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-X2180-1A;
RX MEDLINE=91183402; PubMed=2010462;
RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA Yamasaki M.;
RT "A cytoskeleton-related gene, us01, is required for intracellular
RT protein transport in Saccharomyces cerevisiae.";
RL J. Cell Biol. 113:245-260(1991).
[2]
RN SEQUENCE OF 782-1790 FROM N.A.
RP Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
RA Kendrick K.E.;
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-8 FROM N.A.
RA Bai Y., Symington L.S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
CC COMPLEX.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC ER AND THE GOLGI COMPLEX.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
CC -1- SIMILARITY: BELONGS TO THE VDP/US01/YBL047C FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: X54378; CAA38253.1; -
CC EMBL: L03188; AAB00143.1; -
CC EMBL: U53668; AAB6659.1; -
CC PIR: A38455; A38455.
CC HSSP: P80220; IDIP.
CC SGD: S0002216; US01.
CC InterPro: IPR002017; Spectrin.

```

Query Match 5.3%; Score 187; DB 1; Length 1790;  
 Best Local Similarity 19.9%; Pred. No. 0.016;  
 Matches 150; Conservative 133; Mismatches 273; Indels 198; Gaps 31;

Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.  
 FT DOMAIN 1 724 GLOBULAR HEAD.  
 FT DOMAIN 725 1790 COILED COIL (POTENTIAL).  
 FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).  
 FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.  
 FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).  
 FT CONFLICT 847 847 G -> E (IN REF. 2).  
 FT CONFLICT 924 924 E -> K (IN REF. 2).  
 FT CONFLICT 1253 1253 V -> I (IN REF. 2).  
 FT CONFLICT 1319 1319 I -> V (IN REF. 2).  
 FT CONFLICT 1461 1461 N -> S (IN REF. 2).  
 FT CONFLICT 1581 1581 G -> S (IN REF. 2).  
 FT CONFLICT 1600 1600 I -> V (IN REF. 2).  
 FT CONFLICT 1661 1661 R -> S (IN REF. 2).  
 FT CONFLICT 1772 1772 D -> DEEDDE (IN REF. 2).  
 FT CONFLICT 1772 1772 D -> DEEDDE (IN REF. 2).  
 SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216B9FD4818 CRC64;

Query Match 5.3%; Score 187; DB 1; Length 1790;  
 Best Local Similarity 19.9%; Pred. No. 0.016;  
 Matches 150; Conservative 133; Mismatches 273; Indels 198; Gaps 31;

2 NSGAMRIHSGHFGGCIQVNEKNEKPSLSTKTDNREKSKCPMLGKVFYDLPSTVIS 61  
 978 NESLIKAVESKNESSIOLSNLKNKIDMSOEKEN-----FOIERGSIERN 1023  
 62 -EKLONKIDGGRVEEFLSK-----DISYLISNKKKAPFA--QTIGRISPVSP 108  
 1024 IEQKTIISDLEQTKREIISKSDSKDEYEGISLTKMKLETATNDENVKISLTKI 1083  
 109 ESAYTAETSPHSHSGSPDVCISROKLVKFAIDHDFISNLSIALMSGK 168  
 1084 REELAELEA-----YKMLKELETKLETS-----EALKE--VENE----- 1119  
 169 ILHDDIRYIEOK-----KEELYLKSTSVROGGRVSGAKRTGRLKPPVKED 224  
 1120 -EHLKKEKIQLEKAEETQOOLNSLBNLSEKEDLAKOIKTEQIANKKEQYNEE 1178  
 225 MSQLYRFYQILNMFPIYSTQKCSPPDYD---KSSMOQTQVRLIQTGDKRYGCT 281  
 1179 ISQAND-----ELTSTQOENESITKKNDELBEVAKMSTSEOSNKL-RSEID----- 1226  
 282 SIQQLKKEKKKGYC-----ECCLQKYE--DLETHLLSDQH 315  
 1227 ALNQLIKELKKKNETEASLLSISVSEETVTKIKELQDECNFEKEVESELDKLASED 1286  
 316 RNF-----AQSNOYVVDIVSKLVDFEVEYEDTPKKRKRIYSVGSLSVSAVLKTEQ 371  
 1287 KNSKYLLEQKESKEKEEIDAKTELKIQLEKITNLKAKKESSELSR-----LKTSS 1341  
 372 KEKVELQHSQKDCQDDPTTVEQNFLYKETOETKKLL-----FTSEPIPHPSNELGGL 426  
 1342 EER-----KNAEQQLKELKNEIQIKNOAFKEREKRLNMGSSSTITQYLSKINTLDE 1393  
 427 NEKSNKCSMLSTAEDTIRONTOLPLH-----KNKQECJLISEHTLS----- 470  
 1394 LIRQONNELKAEIDNTSELEKSLSDNDELEKONTKSLQDITLYKAKITRNDK 1453  
 471 -----ENDLELA-----VDHKCNIO-----ASV 490  
 1454 LLSIERNNKDLLESLKQRAAQSAAVEEGKLKEESSKKAELKSKMMKLEST 1513  
 491 HVSDFTSDNGSOPKOKSPTVLFPKADIKELDHLSTFTHDSGLIT--INSOR-----H 542  
 1514 IESNETELKSMSTTIKRSDEKLEQSKSAEEDIKNNLHNKESLISINSSEIDELKSK 1573  
 543 LTVQAK--APFHTPPEEPNEDCFKMDSLPSGKTHRKVLTILGRNKENLFPNAPFDRKT 600  
 1574 LRIEAKGSGLEFVVKOELNNAOEKI-----RINAENVTVL-KSLIEDIERELK-DKOA 1624  
 601 EF-TTOENRICSSPVQSL--LDLFQTSSEKSE 630  
 1625 EIKSNOBEKELLTSRLKLEQLDSTOOKAKRSE 1658

RESULT 2  
 ID AK49\_HUMAN STANDARD; PRT; 3911 AA.  
 AC Q99996; Q9U004; Q9U0H3; Q9Y6V2; Q14869; Q43355; Q94895; Q9Y6B8;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE A KINASE ANCHOR PROTEIN 9 (PROTEIN KINASE A ANCHORING PROTEIN 9)  
 DE (PKA9) (A-KINASE ANCHOR PROTEIN 450 KDA) (AKAP 450) (A-KINASE ANCHOR  
 DE PROTEIN 350 KDA) (AKAP 350) (HGAKAP 350) (AKAP 120 LIKE PROTEIN)  
 DE (HYPERTON PROTEIN) (TOTIAO PROTEIN) (CENTROSOME- AND GOLGI-LOCALIZED  
 DE PKA-ASSOCIATED PROTEIN) (CG-NAP).  
 GN AKAP9 OR AKAP450 OR AKAP350 OR KIAA0803.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 4).  
 RC TISSUE=Brain;  
 RX MEDLINE=98151389; PubMed=9482789;  
 RA Lio J.W., Wyszynski M., Madhavan R., Sealock R., Kim J.U., Sheng M.;  
 RA "Vollao, a novel protein of neuromuscular junction and brain that  
 RA interacts with specific splice variants of NMDA receptor subunit  
 RA NR1.";  
 RL J. Neurosci. 18:2017-2027(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANT GLN-1347 INS.  
 RC MEDLINE=99219864; PubMed=10202149;  
 RX Hiltcak O., Skathnegy E.S., Keyser G., Bornens M., Tasken K.,  
 RA Tamsen T., Oestavik S.;  
 RA Cloning and characterization of a cDNA encoding an A-kinase anchoring  
 RA protein located in the centrosome, AKAP450.";  
 RL EMBO J. 18:1858-1868(1999).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 3).  
 RC TISSUE=Brain;  
 RX MEDLINE=99287934; PubMed=10358086;  
 RA Takahashi M., Shibata H., Shimakawa M., Miyamoto M., Mukai H., Ono Y.;  
 RA "Characterization of a novel giant scaffolding protein, CG-NAP, that  
 RA anchors multiple signaling enzymes to centrosome and the golgi  
 RA apparatus.";  
 RL J. Biol. Chem. 274:17267-17274(1999).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC Kemmer M.A., Dells S., Schwarz U.;  
 RA "Cloning of Hyperion.";  
 RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE OF 323-3911 FROM N.A. (ISOFORM 2).  
 RC TISSUE=gastric parietal cell;  
 RX MEDLINE=99115654; PubMed=9915845;  
 RA Schmidt P.H., Dransfield D.T., Claudio J.O., Hawley R.G.,  
 RA Trotter K.W., Mignem S.L., Goldenting J.R.;  
 RA "AKA950, a multiply spliced, protein kinase A-anchoring protein  
 RA associated with centrosomes.";  
 RL J. Biol. Chem. 274:3055-3066(1999).  
 RN [6]  
 RP SEQUENCE OF 1802-3876 FROM N.A. (ISOFORM 5).  
 RC TISSUE=Lymphoblast;  
 RA Hinds K., Suterer C., Becker M., Hawkins M.;  
 RL Submitted (Jan-1998) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP SEQUENCE OF 2157-3911 FROM N.A. (ISOFORM 6).  
 RC TISSUE=Lung;  
 RA Milgram S.L., Goldenting J.R., Schmidt P.H.;  
 RA "AKA9350: A multiply spliced family of proteins with centrosomal  
 RA association (SEP-1998) to the EMBL/GenBank/DBJ databases.  
 RL Submitted (Sep-1998) to the EMBL/GenBank/DBJ databases.  
 RN [8]  
 RP SEQUENCE OF 2212-3911 FROM N.A. (ISOFORM 2/3).

RC TISSUE-Brain;  
 RX MEDLINE-99087487; PubMed-9872452;  
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,  
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XI.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro.";  
 RL DNA Res. 5:277-286(1998).  
 RN [9]  
 RP SEQUENCE OF 17-1800 FROM N.A.  
 RA Wu X., Graves T., Bradshaw H.;  
 RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: BINDS TO TYPE II REGULATORY SUBUNITS OF PROTEIN KINASE  
 CC A. SCAFFOLDING PROTEIN THAT ASSEMBLES SEVERAL PROTEIN KINASES AND  
 CC PHOSPHATASES ON CYTOSOL AND COLG APPARATUS WHERE PHYSIOLOGICAL  
 CC EVENTS CAN BE REGULATED BY PHOSPHORYLATION STATE OF PROTEIN  
 CC SUBSTRATES. ISOFORM 4/YOTIAO IS ASSOCIATED WITH THE N-METHYL-D-  
 CC ASPARTATE RECEPTOR AND IS SPECIFICALLY FOUND IN THE NEUROMUSCULAR  
 CC JUNCTION (NMJ) AS WELL AS IN NEURONAL SYNAPSES EXPLAINING THAT ITS  
 CC ROLE MAY BE TO ORGANIZE POSTSYNAPTIC SPECIALIZATIONS.  
 CC -1- SUBUNIT: INTERACTS WITH THE REGULATORY REGION OF PROTEIN KINASE N  
 CC (PKN), PROTEIN PHOSPHATASE 2A (PP2A), PROTEIN PHOSPHATASE 1 (PP1)  
 CC AND THE IMMATURE NON-PHOSPHORYLATED FORM OF PKC EPSILON.  
 CC CYTOPLASMIC IN PARIETAL CELLS.  
 CC -1- ALTERNATIVE PRODUCTS: 6 ISOFORMS; 1 (SHOWN HERE), 2, 3/CG-  
 CC NMP, 4/YOTIAO, 5 AND 6/AKAP350; ARE PRODUCED BY ALTERNATIVE  
 CC SPLICING.  
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED. ISOFORM 4/YOTIAO IS HIGHLY  
 CC EXPRESSED IN SKELETAL MUSCLE AND IN PANCREAS.  
 CC -1- DOMAIN: RII BINDING SITE, PREDICTED TO FORM AN AMPHIPATHIC HELIX,  
 CC COULD PARTICIPATE IN PROTEIN-PROTEIN INTERACTIONS WITH A  
 CC COMPLEMENTARY SURFACE ON THE R-SUBUNIT DIMER.  
 CC -1- CAUTION: REF.6 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO TWO  
 CC FRAMESHIFTS IN POSITIONS 3782 AND 3811.  
 CC -1- CAUTION: REF.9 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FOUR  
 CC FRAMESHIFTS IN POSITIONS 29, 1653, 1699 AND 1735.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: AJ131693; CAB40713.1; -  
 DR EMBL: AB019691; BAA78718.1; -  
 DR EMBL: AJ010770; CAA09361.1; -  
 DR EMBL: AF026245; AAB86384.1; -  
 DR EMBL: AF083037; AAD22767.1; -  
 DR EMBL: AC004013; AAB96867.1; ALT\_FRAME.  
 DR EMBL: AF091711; AAD39719.1; -  
 DR EMBL: AB018346; BAA34523.1; -  
 DR EMBL: AC000066; AAC60380.1; ALT\_FRAME.  
 DR MIM: 604001; -  
 KM Coiled coil; Alternative splicing; Polymorphism.  
 FT DOMAIN 2567 PRA-RII SUBUNIT BINDING DOMAIN.  
 FT 164 914  
 FT DOMAIN 944 1022 COILED COIL (POTENTIAL).  
 FT DOMAIN 1100 1185 COILED COIL (POTENTIAL).  
 FT DOMAIN 1253 1280 COILED COIL (POTENTIAL).  
 FT DOMAIN 1336 1392 COILED COIL (POTENTIAL).  
 FT DOMAIN 1434 1459 COILED COIL (POTENTIAL).  
 FT DOMAIN 1585 1659 COILED COIL (POTENTIAL).  
 FT DOMAIN 1857 2455 COILED COIL (POTENTIAL).  
 FT DOMAIN 2544 2561 COILED COIL (POTENTIAL).  
 FT DOMAIN 2603 2776 COILED COIL (POTENTIAL).  
 FT DOMAIN 3065 3092 COILED COIL (POTENTIAL).  
 FT DOMAIN 3124 3470 COILED COIL (POTENTIAL).  
 FT DOMAIN 3587 3689 COILED COIL (POTENTIAL).  
 FT DOMAIN 3726 3730 POLY-LEU.  
 FT DOMAIN 203 292 GLN-RICH.

FT DOMAIN 321 1010 GLU-RICH.  
 FT 1846 2772 GLU-RICH.  
 FT VARSPLIC 17 28  
 FT VARSPLIC 1637 1642  
 FT VARSPLIC 1643 3911 MISSING (IN ISOFORM 2 AND ISOFORM 3).  
 FT VARSPLIC 2175 2182 MISSING (IN ISOFORM 4).  
 FT VARSPLIC 2175 2182 MISSING (IN ISOFORM 4).  
 FT VARSPLIC 2175 2183 MISSING (IN ISOFORM 3).  
 FT VARSPLIC 2895 2907 SADFOKVE -> Q (IN ISOFORM 6).  
 FT VARSPLIC 2895 2907 VGFGRNKFSTLC -> GSTPELAHSAVOTREICSS  
 FT VARSPLIC 2895 2907 (IN ISOFORM 2, ISOFORM 3 AND ISOFORM 6).  
 FT VARSPLIC 2895 2907 MISSING (IN ISOFORM 5).  
 FT VARSPLIC 3901 3911 STGFHAGMR -> ALSITTSWOHSARPALEFELISH  
 FT VARSPLIC 3901 3911 SLG (IN ISOFORM 6).  
 FT VARIANT 1347 1347 K -> KO.  
 FT CONFLICT 76 76 E -> Q (IN REF. 3).  
 FT CONFLICT 475 475 M -> I (IN REF. 3).  
 FT CONFLICT 554 554 E -> G (IN REF. 3).  
 FT CONFLICT 638 638 R -> S (IN REF. 3).  
 FT CONFLICT 663 663 N -> S (IN REF. 3).  
 FT CONFLICT 913 913 H -> N (IN REF. 3).  
 FT CONFLICT 956 956 K -> N (IN REF. 3).  
 FT CONFLICT 980 982 OKH -> PKP (IN REF. 1 AND 2).  
 FT CONFLICT 997 997 Q -> P (IN REF. 1 AND 2).  
 FT CONFLICT 1001 1001 Q -> P (IN REF. 1 AND 2).  
 FT CONFLICT 1020 1020 N -> D (IN REF. 3).  
 FT CONFLICT 1028 1028 V -> E (IN REF. 3).  
 FT CONFLICT 1626 1626 R -> P (IN REF. 1 AND 2).  
 FT CONFLICT 1703 1703 N -> T (IN REF. 3).  
 FT CONFLICT 1707 1707 V -> G (IN REF. 3).  
 FT CONFLICT 1802 1803 MISSING (IN REF. 5).  
 FT CONFLICT 1843 1843 A -> P (IN REF. 3).  
 FT CONFLICT 1956 1956 I -> V (IN REF. 3).  
 FT CONFLICT 2027 2027 V -> D (IN REF. 5).  
 FT CONFLICT 2157 2158 EI -> HE (IN REF. 7).  
 FT CONFLICT 2159 2159 E -> V (IN REF. 3).  
 FT CONFLICT 2514 2514 L -> R (IN REF. 3).  
 FT CONFLICT 2851 2851 I -> N (IN REF. 8).  
 FT CONFLICT 2957 2957 E -> D (IN REF. 3).  
 FT CONFLICT 2983 2983 P -> S (IN REF. 3, 7 AND 8).  
 FT CONFLICT 3087 3087 Q -> H (IN REF. 3).  
 FT CONFLICT 3218 3218 Q -> H (IN REF. 3).  
 FT CONFLICT 3307 3309 ESE -> OSO (IN REF. 3).  
 FT CONFLICT 3751 3751 P -> A (IN REF. 3).  
 FT CONFLICT 3833 3833 T -> S (IN REF. 3).  
 FT SEQUENCE 3911 AA; 453664 MW; 3FB1CB1819B4/AA CRC64;

Query Match 5.1%; Score 178; DB 1; Length 3911;  
 Best Local Similarity 21.1%; Pred. No. 0.13;  
 Matches 115; Conservative 104; Mismatches 184; Indels 142; Gaps 27;  
 QY 179 IEQKKRELYLLKKSSTVADGKRGVSGAOKTRTGRLLKPFVVEDMSQLYRPFYLQITN 238  
 DB 351 IEEKKTELELKKOKLTADKLGLGEQIVQKNQ--EIKRMKLELINSKOKESOSSEIKO 408  
 QY 239 MEFITYSIQ---PCSPEDVNDKPSMKOKOTQV---LRITGSGDKYGGSTIDL 285  
 DB 409 IMGVEELQKRNHKSOFETDVGORMEOTQRLQRLAELD-EMVGOQIVYKQELIRQ 467  
 QY 286 ---OLKEKK--KKGCECLOKEDLETH-----LSPQHNHFNQSNQ----- 323  
 DB 468 HMAQMEKTRKRGKEMENALRSTNYVNEQDIKLMAVAINELINKLQIDTNSQKEKKE 527  
 QY 334 -----YQVVDIVSKLVF--DFVEYBKDPYKKRIKY-----SVGSLSPVSASV 365  
 DB 528 LGLILEKCALQRLQEDLYVELSFRBQIQRARQTIAGESKLNENHKSLSLVEEDLKAEI 587  
 QY 366 IAKTQKEKVELQHTISQKDCQEDDTT-----VKQENFL---YKETQEPF-----KK 408  
 DB 588 VSASESRKELELKH-----EAEVTNFKIKLEMLEKEKNALVLDMAESQAELERLRQ 640  
 QY 409 LFISE-----PIPHPSN-----ELRGINEMSNKSCSLSTAEDDI 444

DB 641 LIFSHEERLEKLEDEIEHRIINIKLNDIGIHYKQOIGLONEMSKIEITMOPEKNDL 700  
 OY 445 RQNTFOLPLH-----KMKOCILD--ISEHTLSENDLE-ELRV-----DHYKNCIOASVHS 493  
 DB 701 ITRKONOLLEISKIKDLOOSLVNSKSEMTQJINELQKEIILROEKEKGTLEOVEL 760  
 OY 494 DESPDNGSGOPKQSPVLFPAKDKERKDLHSIFTHSGCLITNS--SOEHLTVQAKAPF 551  
 DB 761 QLKTELLEKQMKKEK-----DLQEK-----PAQLEANSILKDEKKTLEDMLKI 805  
 OY 552 HTPEPEENEDDFNMOSLPS---GKIHKVKIILGN---RKENLEPNAPFDDK-RTEFI 603  
 DB 806 HTPVSOEERLIF--LDSIKSKSDSWKEKEIIEIIFENEDLKQOCIOULNEIEKQKRNFTS 863  
 OY 604 TOEN 608  
 DB 864 PAEN 868  
 RESULT 3  
 REST\_CHICK STANDARD; PRT: 1433 AA.  
 ID 042184: 042228; 057563; 057564;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE RESLIN (CYTOPLASMIC LINKER PROTEIN-170) (CLIP-170).  
 GN RSN.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 CC NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98137792; PubMed=9469933;  
 RA Gripatic L., Volosky J.M., Keller T.C. III;  
 RT Cloning and expression of chicken CLIP-170 and reslin isoforms.\*;  
 RL Gene 206.195-206(1998).  
 [2]  
 RP SEQUENCE OF 17-1139 FROM N.A. (CLIP-170(11) AND CLIP-170(11+35)).  
 RA TISSUE=Pectoralis muscle; III;  
 RC Gripatic L., Keller T.C. III;  
 RT Identification and expression of two novel CLIP-170/Reslin isoforms  
 expressed predominantly in muscle.\*;  
 RL Submitted (Feb-1998) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN  
 THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE  
 CYTOSKELETON (BY SIMILARITY).  
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 4 ISOFORMS ARE PRODUCED BY  
 ALTERNATIVE SPLICING.  
 CC -1- SIMILARITY: CONTAINS 2 CAP-GLY DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-slb.ch/announce/  
 or send an email to jlicense@isb-slb.ch).  
 CC -----  
 CC EMBL: AF014012; AAC60344.1; -  
 DR EMBL: AF020764; AAC60345.1; -  
 DR EMBL: AF045650; AAC03547.1; -  
 DR EMBL: AF045651; AAC03548.1; -  
 DR InterPro: IPR000938; CAP-GLY.  
 DR InterPro: IPR001878; zn1\_CCHC.  
 DR Pfam: PF01302; CAP\_GLY; 2.  
 DR SMART: SM00343; znF\_C2HC; 1.  
 DR PROSITE: PS00845; CAP\_GLY\_1; 2.  
 DR Cytoskeleton; Microtubules; Colled coil; Alternative splicing.  
 FT DOMAIN 79 121 CAP-GLY 1.

FT DOMAIN 144 207 SER-RICH.  
 FT DOMAIN 235 277 CAP-GLY 2.  
 FT DOMAIN 305 332 SER-RICH.  
 FT DOMAIN 351 353 COILED COIL (POTENTIAL).  
 FT DOMAIN 1414 1427 CCHC-BOX.  
 FT VASPLIC 458 492 MISSING (IN SHORT ISOFORM).  
 FT VASPLIC 458 492 TOTLEHARKLEQSLFETKAKDLQRELEDT ->  
 FT VASPLIC 803 803 RKRQISEDEN (IN ISOFORM CLIP-170(11)).  
 FT VASPLIC 803 803 T -> RKRQISEDEPT (IN ISOFORM CLIP-  
 FT VASPLIC 458 458 170(11+35)).  
 FT CONFLICT 309 309 K -> R (IN REF. 2; AAC03547).  
 FT CONFLICT 440 440 E -> V (IN REF. 2; AAC03548).  
 FT SEQUENCE 1433 AA; 161026 MW; 5631CE8683498E23 CRC64;

Query Match 4.7%; Score 165; DB:1; Length 1433:

Best Local Similarity 18.7%; Pred No 0.17; Mismatches 144; Gaps 32;  
 Matches 135; Conservative 139;

OY 11 KGFHGGIOYKN-----EKNRPSLKSIKTDNRPKSKCPKMGKPYFL-----DLPS 57  
 DB 771 KANSEKLEIOLKLEQLOAKERQIONLETEKVSN-----LTKELQKEKQLDEKKNLSA 825  
 OY 58 VT-ISEKLOKDK-----DIGRVEEFLSKDISYILSKKKAAPQTLGRISPPV 106  
 DB 826 VNOVKOSLEKEIOLKEKETSADVDAENAOAQMOTINKL--NOREEQPALMSELQOLK 883  
 OY 107 SPESAYTAETTSPPSHDSSSEKSPDTVCLSRKLLVERKAKIDHETPSNLSLSWGC 166  
 DB 884 S--NLVTMET-----KLKERE-----EREQOLTEAKVKLENDIA--ELMSSGDS 925  
 OY 167 VKLIHIDIIYYIEQKKKELYL-LKSSSTSVROGKKRVSGAOKRTGTRKKPKVYEDM 225  
 DB 926 AQLKMKNDLEKLEKOLEIOLLEITKANERKAVOLAKNVEQTAQKEOSQOETLTHQEL 985  
 OY 226 SOLRPPYLOLTMPRPINYSIOKPCSPDVDRPSSMOKOTQVRLRIQT-----DGGKY 278  
 DB 986 KKMDD--QLTDM-----KKOMETSQNYKDLQAKYKETSMTIKKIDADIK 1029  
 OY 279 GGTSTIOLKKEKKKKKGCYCCLOKYEDELTEHLSEGRNFAQNGOVQVIVSLYVDF 338  
 DB 1030 GFKNLDAEALK-----AAQKNDELETO-AEELKKQAQAKKRAEVLQTM--EK 1081  
 OY 339 VEYKOTPKKKRKIKYVSVLSVSAVLAKTEQKEKELQHSIQKDCOPDQTYEQNFL 398  
 DB 1082 VTKEDAIHQEKTE--TLASLENS--KOTNEKLQNELMLKONNKLKNEELIKSKEL 1135  
 OY 399 ---YKETOETEKLFISEPIHPSPNELRGINKMSKCSMLSTAAEDDIRQNTOLPLHK 455  
 DB 1136 NLENKRYEELKKEFEFLAAQKSOQLALQDEENVKLABELGRSDEVTs-----HQ 1188  
 OY 456 NKOCILDISENTLSENDLELR-----VDHYKNCIOASVHSDS--IDNSGQ 503  
 DB 1189 KLEP-----ERSVLLNNOLLEMKKRESLTKKEIDEERASLOKS--ISDIALITQKDEL 1240  
 OY 504 PKQSDVLPFPAKDKERKDLHSIF-THDSGLITINSQEHITVQAKAPFPPPEPNCD 562  
 DB 1241 EKLNETTVYRGNGENAKTIOGVVTLSEDKLKEEKVKNLEOKLAKSEDFLYVTS-- 1297  
 OY 563 FKNDSLPSCKIRKVKIILGNRKNLEPNAEF-----DKRFEFTQENRNCSSP 614  
 DB 1298 -----PSGDIAN--LLODESADKQOEIDFLNSVIVDQRRNEELNLKIQMCEAA 1347  
 OY 615 VQSLDLDFQSEKSEKSEFLGFTSYTEK--SGICNVLDIWEENSNDNLTAFFSSPSTFT 672  
 DB 1348 LNGEE--ETINVDSEEBLSKTRFLPCDLCGCFDLDHEDCPT-QAQMLEPPHSHYH 1404  
 OY 673 G 673  
 DB 1405 G 1405



GN RSN.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Peripheral blood monocytes;  
MEDLINE=92289675; PubMed=1600942;  
RA Balke G., Delaite J., Bruggen J., Richener H., Asselbergs F.A.M.,  
Celletti N., Sorg C., Orlan K., Tarsasy L., Wiesendanger W.,  
de Wolf-Petersen C., Shilman R.;  
RT Result: a novel intermediate filament-associated protein highly  
expressed in the Reed-Sternberg cells of Hodgkin's disease.;  
RL EMO J. 11:2103-2113(1992).  
[2]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=92405160; PubMed=1356075;  
R Pierre P., Scheel J., Rickard J.E., Kreis T.E.;  
RT "CIP-170 links endocytic vesicles to microtubules.";  
Cell 70:887-900(1992).  
CC -1- FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN  
THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE  
CYTOSKELETON.  
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A  
SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE REED-STERNBERG CELLS  
OF HODGKIN'S DISEASE.  
CC -1- SIMILARITY: CONTAINS 2 CAP-GLY DOMAINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isp.slb.ch/announce/>  
or send an email to [license@isp.slb.ch](mailto:license@isp.slb.ch)).  
CC  
DR EMBL; X64838; CA646050.1; -  
DR EMBL; M97501; AA653693.1; -  
DR PIR; S22695; S22695.  
DR MIM; 179838; -  
DR InterPro; IPR000938; CAP-GLY.  
DR InterPro; IPR001878; znf\_CCHC.  
DR Pfam; PF01302; CAP-GLY; 2.  
DR SMART; SM00343; znf\_C2HC; 1.  
DR PROSITE; PS00845; CAP-GLY\_1; 2.  
KW Cytoskeleton; Microtubules; Coiled coil; Alternative splicing.  
FT DOMAIN 60 125 CAP-GLY 1.  
FT DOMAIN 143 204 SER-RICH.  
FT DOMAIN 214 279 CAP-GLY 2.  
FT DOMAIN 304 331 SER-RICH.  
FT DOMAIN 350 1342 COILED COIL (POTENTIAL).  
FT DOMAIN 1408 1421 CCHC-BOX.  
FT VARSPIC 457 491 MISSING (IN SHORT ISOFORM).  
FT COMPLET 1069 1069 D -> E (IN REP. 2).  
SQ SEQUENCE 1427 AA; 160989 MW; 0A6F166D94254EB CRC64;

DB 871 H0KE0FNMSSDLEKRENLADMEAKFREK----- 902  
OY 180 E0KKKEYLKKSSTSVBDGKRVG-SGAOKRTGT--RLKRPFKVVDMSQLYRPYLO 235  
DB 903 EREBOLIKAKKEKLENDIAEIMKMGDSSQLTKNDELRLKE-----RDVEL-----OLK 953  
OY 236 LT-----NMPFINSIOKPCSPEDVDKPSMOKOTVAKLRIGDGDGKGTSGIOLKEKK 291  
DB 954 LTKANENASPLQKSIEDMTVAEQSQEAAKKEHEKKELER-----KLSDLEKKMET 1006  
OY 292 KKGCGECQCKYEDLETHLSBOHRNFAOSNOYOVD-----DIVSLYVDPEYE 342  
DB 1007 SHNCOELAKRYE-RATSETKTKHEILQNLQKTLDPTEDKLGARENSGLQLELELR 1065  
OY 343 KDTPKKRIKISVSGLSFVSASVYLKTTQKEKVELDHISQKQCDQDDDTVEKQFLKET 402  
DB 1066 KQADKAAQOTADDAQIMQKTEKTEETLASLEDTKQTKAKLQNELDLKENVKE 1125  
OY 403 QETKRLFT-SEPIPHSPNELRGLNEKMSKCSMTVAE-----DRIKQFTQLPLH 454  
DB 1126 LNKSKELLVYENQKMEEFKELETLKQAQAKSOOLALQENVKLAELGSDREVTS 1185  
OY 455 KKNQECILDISEHTLSENDLELR-----VDHYKNQASVHV-SDFSTDNGSQP 504  
DB 1186 QKLEE-----ERSVYNQLLEMKKRESKFIKQADEKASLQKSITSITALLTEKDAELE 1239  
OY 505 KQKSDTVLPFAKDLKEKDLHSIF-THDSGLITINSQEHLYQAKAPFHTPEEPNCD 563  
DB 1240 KLRREVTVLGENASAKSLHSVQTLSDSKVLELKVKYKLELOLK-----EN 1286  
OY 564 KNDSDLSGIRHKVYKILGRNRKENLEPNAEPDKRT-----EFTT-----QENRIRCS 612  
DB 1287 KRLSSSG-----NDTQADEDERAEOESQIDFLNSVYDQKRNQDLK 1330  
OY 613 SPVQSL-----LDLPQTSKEKEEFLGFTSYEKSGICNVYDIWEENSNNLLT- 660  
DB 1331 MKVYMSEALNNGDNLNNDSDQEQ-----SKKKPRLDICDQCDLHDREDCQTQ 1385  
OY 661 -AFSSPSPTFTG 673  
DB 1386 AQMSDEPPSTHNG 1399  
RESULT 6  
ID CENF\_HUMAN STANDARD; PRT; 3210 AA.  
AC P49454; Q13246; Q13171;  
DR 01-FEB-1996 (Rel. 33, Created)  
DR 01-FEB-1996 (Rel. 33, Last sequence update)  
DR 20-AUG-2001 (Rel. 40, Last annotation update)  
DE CENP-F KINECHORE PROTEIN (CENTROMERE PROTEIN F) (MITOSIN) (AH  
ANTIGEN).  
GN CENP-F.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Breast carcinoma;  
MEDLINE=95348173; PubMed=7542657;  
RA Liao H., Winkler R.J., Mack G., Ratner J.B., Yen T.J.;  
RT "CENP-F is a protein of the nuclear matrix that assembles onto  
kinetochores at late G2 and is rapidly degraded after mitosis.";  
J. Cell Biol. 130:507-518(1995).  
[2]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=95379848; PubMed=7651420;  
R Zhu X., Mancini M.A., Chang K.-H., Liu C.-Y., Chen C.-F., Shan B.,  
Jones D., Yang-Feng T.L., Lee W.-H.;  
RT "Characterization of a novel 350-kilodalton nuclear phosphoprotein  
that is specifically involved in mitotic-phase progression.";



CC -1- SIMILARITY: TO S.POMBE SPAC20G8.09C AND AN A.AMBISEXUALIS  
CC HYPOTHETICAL PROTEIN (AC P54008).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL: 246843; CAAB6893.1; -  
CC DR EMBL: 271408; CA96014.1; -  
CC DR SGD: S0005076; YNL132W.  
CC Hypothetical protein: ATP-binding.  
CC NP\_BIND 283 290 ATP (POTENTIAL).  
CC SEQUENCE 1056 AA; 119347 MW; 76721ED0867ED618 CRC64;

Query Match 4.4%; Score 155.5; DB 1; Length 1056;  
Best Local Similarity 21.0%; Pred. No. 0.37; Indels 235; Gaps 37;  
Matches 151; Conservative 98; Mismatches 235; Indels 235; Gaps 37;  
19 QVKNKRNPSLSLKT---DNRPKSKCPLMGKVFYLDLPSTVISE-KLOKDIKDL- 71  
429 QLRNONNTSGESOTGVVSRDKKERDS-----HLMSQROLRETSIDELRVAP 478  
72 GGRVEEPLSKDISYLSNKKKAKFA-----OTLGRISPEPS- 107  
479 GPRIKMLNKLICDVTLLKNPFRATRGPHPSQCNLFVVRDITFSGHSENFLEKMM 538  
108 -----PESAVTAETTPSPH-----DSSSKSPDYTC----- 135  
539 ALVSSHKKNPNQIMSDAPAHKFLVLPIDPKDGG-RIPDPICVQIALGEISK 596  
136 -----LSRG-----LLVERAKHDHDFPSMS-----ILSN-----ALMGVYLII 172  
597 ESVRNSLSRGRAGDILPILISQOFODEFASLSGAIIVAIKIANPEYASMGSGAIEL 656  
173 DDIRYIQQKKELYLLKSSSVKPGC---KRVSGAQKTRGRLLKPKVKEDESOIX 229  
657 --LRDYFEGKFTDM-----SEVPRKDYSIKRV--SKELAKTMLKMD--VKLRDKTL- 705  
230 RPYLDLTNMP-----FTNSIQKPCSPEDVDRPSSMOKOTQVLRITQDCKYTG-GT 281  
706 PELLKLTSEQPHYLHYGVSYGLTQSLHF--WKNSF---VPEYLR-QTANDLTGEHT 759  
282 STIOLAKKKKKGCCCLQKYEDLETHLSEQHRNFASQNOYVVDIVSKLVFDEVEY 341  
760 CVMALNVLGRESNMVLEFAKDFPKRRLSLSYDFHKFTAVOALSVIES--SKRAODLSDD 817  
342 EKDTPEKK-----KRI-KYS-----VGSISPSASAVLKKTQCKREVEL 377  
818 EKHDKKELRTHLDIFSPDLKRLDSYNNLLDYHVGIMIPMLALLYFGDKMGDSVKL 877  
378 QHISQK-----DCQEDPTVEKQNPFLVKPEOEKLLPISLPHPSNELRGLENEM 430  
878 SSVQSAIILAIQLOKRIIDITAKELN-----PSNQIITAFAKI 916  
431 SKKSGMLSTAEEDIRQNFQPLAKKQCCIDISEHLSLSENDLEIKLRVHYKNCIOASY 490  
917 MKRMS-----QYFROL-LSQSTIEELPRIKODIALEKDEEIK-----NINAAE 959  
491 HVSDESTD--NSGQPKQSDVLEFPAKDLKEKDLHSIFTHDSGLITNS-----SOEHL 543  
960 ALDQEEDELEFAGSEAVQ-----AMREKQKELINSLNDKYAINDNSEMAEASQKSL 1011  
544 TVQAKAPRHTPEEPNECDPKNMDSLPSGK-----IHKKYKIIIGRRKRENEPN 593  
1012 EIAAKA-----KGVASLTKGKRTTEKADIYFOEMKAMKRPKRKSKAAN 1056

RESULT 8

CENE\_HUMAN STANDARD; PRT; 2663 AA.  
AC 002224;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE CENTROMERIC PROTEIN E (CENP-E PROTEIN).  
GN CENP-E.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP MEDLINE=93024922; PubMed=1406971;  
RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.,  
RT "CENP-E is a putative kinetochore motor that accumulates just before  
RT mitosis.";  
RL Nature 359:536-539(1992).  
RN [2]  
RP CHARACTERIZATION.  
RA MEDLINE=9516755; PubMed=7889940;  
RT Thowar D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;  
RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed  
RT microtubule motor.";  
RL EMBO J. 14:918-926(1995).  
RN [3]  
RP CHARACTERIZATION.  
RA MEDLINE=96437347; PubMed=9763420;  
RT Chan G.K.T., Schaar B.T., Yen T.J.;  
RT "Characterization of the kinetochore binding domain of CENP-E reveals  
RT interactions with the kinetochore proteins CENP-F and hBUBR1.";  
RL J. Cell Biol. 143:49-63(1998).  
CC -1- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE  
CC KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE  
CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT  
CC AND/OR SPINDLE ELONGATION.  
CC -1- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.  
CC -1- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING  
CC CONGRESSION. RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS  
CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.  
CC  
CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL: 215005; CAA78727.1; -  
CC DR PIR: S28261; S28261.  
CC DR HSSP: P03069; 1211.  
CC DR MM: 117143; -  
CC DR InterPro: IPR001752; Kinesin.  
CC DR Pfam: PF00225; Kinesin; 1.  
CC DR PRINTS: PRO0380; KINESINHEAVY.  
CC DR SMART: SM00129; KISC; 1.  
CC DR PROSITE: PS00411; KINESIN MOTOR DOMAIN1; 1.  
CC DR PROSITE: PS50067; KINESIN MOTOR DOMAIN2; 1.  
CC KW Motor protein; Cell division; ATP-binding; Colled coil; Mitosis;  
CC Cell cycle; Centromere.  
CC KM DOMAIN 1 335 MECHANOCHEMICAL (MOTOR).  
CC FT DOMAIN 336 2471 COILED COIL (POTENTIAL).  
CC FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).  
CC FT NP\_BIND 86 93 ATP (BI SIMILARITY).  
CC SEQUENCE 2663 AA; 312087 MW; CEFCL380C8C8C88 CRC64;

Query Match 4.4%; Score 154; DB 1; Length 2663;  
Best Local Similarity 18.2%; Pred. No. 1.5;  
Matches 147; Conservative 117; Mismatches 243; Indels 300; Gaps 37;



```

OY 18 IQVKNKRNPSLSKLTQNR-----PEKSKCPYLMGKVVYLDLPSTVTSKIQKIDKDG 72
DB 1379 IQESQSKQDSINMKEKNETTKIYSEMEQFQPKPSALRIEMIGLSKRLQESHDEM- 1437
OY 73 GRVEEFLSKDISYLSINKEAKFAOTLGRISVPSPESAVTAETTSPPHSHDSSPKSPD 132
DB 1438 -----KSYAKEKDDLRLOEVLOSQSOLKENI----- 1465
OY 133 TVCLSRGKLVEKAIKDHDFPSNSILSNALSMGVKILH-----IDIRYYIEOKK 183
DB 1466 -----KEIVAKHLETEE-----ELKVAACJLKEQEEFINELRVNLSERE 1504
OY 184 KELYLKSSSVSRV-----CGKRVGSAQOTFRGRLKPPRYKVEDMS- 226
DB 1505 TEISIQOLEKINIKLONKIOEIEYEKEQNLKQISEVEQVNVNELKQKERRKADSL 1564
OY 227 QLYRPFYQLTMPPIVINSIQKCPSPFDV-DKPPSMOKOTQVRLRIOTDGYGCTSIQL 285
DB 1565 OSIESKMLETN-----RLQESQEEIQMIKEKEMKRVQALQIERDQLENKKEIYA 1618
OY 286 QLKERRKRGY-----CECCLOREYEDLETHLSEQ-----HRN 317
DB 1619 KMKESQEKVEYQPLKMTAVNEQEKMEIEHLK-EQEFQKLMLENIEFENIRLTQILHEN 1677
OY 318 FQGSNOY-QVVDIYKSLVDFVEYEKDPKRRKRIYSGSLSPSASVLTKEQEKYE 376
DB 1678 LEEMSVTKERDILRS--VEETLKVERDOLK-----LREITRLERKQEEIK 1724
OY 377 LOHI-----SQRDCEBDTTVKQONFL-----YK 400
DB 1725 IVHMLKHEQETIDKLRIGVSEKTEISMQKDLHSNALKAQDIOEELRIAMHLK 1784
OY 401 ETOETEKILL-FISEPIPHPSN--ELRGLNEMSKNSKCS-----MLSTAEDDIRQ--- 446
DB 1785 EEOETIDKLRIGVSEKTEKLSNMQDLHNSNAKIQEIKELKANEHQLITLKNQVNETQK 1844
OY 447 -----NFTQLPHKNKQEC-----IDLI 464
DB 1845 KYSEMEQKQIKQKDSLTSLKLEIENLDAQ-ELHENEEMKSVKERNDLRVEETLKL 1903
OY 465 SEHTLSEN-----DLE--ELR-----VDHYKCNII-OASVHVSDFSTDSNG 501
DB 1904 EBDQKESIOETKARDLEIQOELKTRAMLSKHEKTEVDLREKISKITQIDQICDLK 1963
OY 502 SOPKQKSDTVLEFPANDLEKDIHSIFTHDSGLITINSQEHLT--VOAKAPFHTPEEPN 559
DB 1964 SK-----DELQKKIQEIQELQKELQLLRKED-----VMSHKIKINEMOLKQF-----EPN 2009
OY 560 ---EEDFKMMDLPSGKIH--RKVKILG-----RNKENLEPNAEPDKRTIEFT----- 603
DB 2010 YLCKCEMDNFQI--TKKLHSELEERITVAKERDEDLRIRIKESIKM-----ERDQFIATLRE 2062
OY 604 -----TOENRNICSSPQSLD 620
DB 2063 MIARDQRNHQVPEKRLSDSQOHLME 2089

```

```

RP SEQUENCE FROM N.A. (ISOFORMS A AND B).
RC TISSUE-Brain;
RX PubMed-11118617;
RA Caricasole A., Sala C., Roncarati R., Formentti E., Terstappen G.C.;
RT Cloning and characterization of the human phosphoinositide-specific
RT phospholipase C-beta 1 (PLCbetal).";
RL Biochim. Biophys. Acta 1517:63-72(2000).
RN [2]
RN SEQUENCE FROM N.A. (ISOFORM A).
RC TISSUE-Brain;
RX MEDLINE-20225428; PubMed-10760467;
RA Peruzzi D., Calabrese G., Faenza I., Manzoli L., Matteucci A.,
RA Gianfrancesco F., Billi A.M., Stuppia L., Palka G., Cocco L.;
RT Identification and chromosomal localisation by fluorescence in situ
RT hybridisation of human gene of phosphoinositide-specific phospholipase
RL C beta 1.";
RL Biochim. Biophys. Acta 1484:175-182(2000).
RN [3]
RN SEQUENCE OF 261-1216 FROM N.A. (ISOFORM B).
RC TISSUE-Testis;
RA Koehler K., Beyer A., Mewes H.-W., Gassenhuber J., Wilmann S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RN SEQUENCE OF 519-1216 FROM N.A. (ISOFORM A).
RC TISSUE-Brain;
RX MEDLINE-98290545; PubMed-9628581;
RA Nagase T., Ishikawa K.-I., MiyaJima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:31-39(1998).
CC -1- FUNCTION: THE PRODUCTION OF THE SECOND MESSENGER MOLECULES
CC DIACYLGLYCEROL (DAG) AND INOSITOL 1,4,5-TRISPHOSPHATE (IP3) IS
CC MEDIATED BY ACTIVATED PHOSPHATIDYLINOSITOL-SPECIFIC PHOSPHOLIPASE
CC C ENZYMES.
CC -1- CATALYTIC ACTIVITY: 1-PHOSPHATIDYL-D-MYO-INOSITOL 4,5-BISPHOSPHATE
CC + H(2)O -> D-MYO-INOSITOL 1,4,5-TRIPHOSPHATE + DIACYLGLYCEROL.
CC -1- COFACTOR: REQUIRES CALCIUM.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A (SHOWN HERE), AND B; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- MISCELLANEOUS: THE RECEPTOR-MEDIATED ACTIVATION OF PLC-BETA 1 IS
CC MEDIATED BY TWO G-PROTEIN ALPHA SUBUNITS, ALPHA-O AND ALPHA-11.
CC -1- SIMILARITY: DOMAINS X AND Y ARE CONSERVED IN DIFFERENT FORMS
CC OF PLC AND ARE ESSENTIAL FOR CATALYTIC ACTIVITY.
CC -1- SIMILARITY: CONTAINS 1 C2 DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AJ278313; CAB98142.1; -
DR EMBL: AJ278314; CAB98143.1; -
DR EMBL: AY004175; AAF86613.1; -
DR EMBL: AL137267; CAB70666.1; -
DR EMBL: AB011153; BAA25507.1; -
DR HSSP: P10688; 10AS.
DR InterPro: IPR000008; C2.
DR InterPro: IPR001192; PL_PLC.
DR InterPro: IPR000909; PL_PLC_X.
DR InterPro: IPR001711; PL_PLC_Y.
DR Pfam: PF00168; C2; 1.
DR PRINTS: PR00390; PHPLIPASEC.
DR PRODOM: PD001202; PL_PLC_Y; 1.
DR SMART: SM00239; C2; 1.
DR SMART: SM00148; PLCXc; 1.
DR SMART: SM00149; PLCYc; 1.
DR PROSITE: PS50004; C2_DOMAIN_2; 1.
DR PROSITE: PS50007; PIPLC_X_DOMAIN; 1.

```



```

Db 392 HEMGOFAPFISODLTDITDOR-AKDIDL-----KETNSDLIKSTT-VDSONLEYKMKDR 444
Qy 118 SP--HPSHDGS-----SPKSPDYVCLSRGKLLVEAKIXDHDFIPNSIISNALSMGVKILH 171
Db 445 SKLHDSLELAKELKSPFSLST-----QDSLNHELENLKTVEKYLQSMSEMI- 492
Qy 172 IDDIRYIEQKKKELLYLKKSSTSVROGKRVGSGAQ-----KRTGKLKPF 219
Db 493 IPKLNOKLEEKNEMMIJLENOIEKQDRIMKTNOADLYAKGLIKKSINIKRLDELQKIT 552
Qy 220 VKVEDMSQLYPEFYL-----OLTNMPFIN 243
Db 553 EKLQDSIRIOVFPLTQEFQADLEMDQKLFINMOKNIAINNMHMLDRRYTALINL 612
Qy 244 YSIQPCPS-----PFD--VDKPSMOKOTQVKRLIOTDGRYSGTSIQ 284
Db 613 NTIEKIDQDNQKSEKVIQQLSENLPEDCTIDEYNDVLEFELSYKTALENLKMHTLE 672
Qy 285 LQ---LAKKKKKGCECLOKVEDETHLLSEQHNFASQNOYQVVDIVSKLVDPVEY 341
Db 673 FNRRALAEIARDSCCYLCSRKFEN-----ESFESKLLQELKTKTDANFERT 718
Qy 342 EKDPKKRR-----IKYSGSLSPVSASY--LKTEQKKEVLOHISOK--DCQEDD 389
Db 719 LKDYQNKKEVLIHSLRLLEKHITTLNSINEKIDNSQCKLERAKETKTSKSLDELDVDS 778
Qy 390 TTVKQNFLYKETQETEKLLFISEPIPHSPNELRGL--NEKMSNKCMLSTAD-- 442
Db 779 TKLKE-----KELASEIRPLI--EKFTYLEKELKLENSKTSIEELIYTSDEGIQT 832
Qy 443 -----DIRQNFOTPLKKNQ-----ECILDISHTISENLEELRDH 481
Db 833 VDELQDQQRKNDLSRELKJISDQMEKDEKVRNSRMINIKERELVSEISSLTQ- 891
Qy 482 KYCNIGASVHVSDFSTDSGSGOPKQSDTVLEPAKDLKEDLHSTFHDGSLITNSQSE 541
Db 892 -KONIDDSIRSKREINIDISRVKELERIT--SLKMK-----KDEASVLDKAKN 939
Qy 542 HLTVQAKAP-----FHTPPEPNECDPKNMDSLPSGRIHRKVKITILGNRKE 588
Db 940 ERDIOVRKQKQTVADINHLIDRFQITVEVVDVFEAKGFDELQT-----TIK-ELELNKQ 993
Qy 589 NLEPNAEFDKRTFTTQEBNRICSS--PVQSLDLFQTSSEKSEFLGFTSYEKGICNV 646
Db 994 MLELKEQDLKLSNEVNEERKLADSNENNEKNLQMLIELKSQLQIHSEISRLDVQNA 1053
Qy 647 ---LDIWEENSNDLTAEP 662
Db 1054 EAERDKYQEE-SLRLLTRP 1071

RESULT 11
MY9B_RAT MY9B_RAT STANDARD: PRT: 1980 AA.
AC 063358;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE MYOSIN IXB (UNCONVENTIONAL MYOSIN-9B).
GN MYO9B OR MYR5.
OS Rattus norvegicus (Rat.).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain stem, and Spinal cord;
RX MEDLINE=95188874; PubMed=7882973;
RA Reinhard J., Scheel A.A., Diekmann D., Hall A., Ruppert C.,
RA Baehtler M.;
RT "A novel type of myosin implicated in signalling by rho family
RT GTPases.";
```

```

RL EMBL J. 14:697-704(1995).
CC -I- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE
CC ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.
CC MAY BE INVOLVED IN THE REMODELING OF THE ACTIN CYTOSKELETON. BINDS
CC ACTIN WITH HIGH AFFINITY BOTH IN THE ABSENCE AND PRESENCE OF ATP
CC AND ITS MECHANOCHEMICAL ACTIVITY IS INHIBITED BY CALCIUM IONS.
CC ALSO ACTS AS A GTPASE ACTIVATING PROTEIN ON RHO.
CC -I- SUBCELLULAR LOCATION: IN UNDIFFERENTIATED CELLS COLOCALIZES WITH
CC F-ACTIN IN THE CELL PERIPHERY WHILE IN DIFFERENTIATED CELLS ITS
CC LOCALIZATION IS CYTOPLASMIC WITH THE HIGHEST LEVELS IN THE
CC PERINUCLEAR REGION (BY SIMILARITY).
CC -I- TISSUE SPECIFICITY: EXPRESSED IN TESTIS, LUNG, THYMUS, BRAIN,
CC LIVER, AND SPLEEN.
CC -I- SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHOSPHO-ESTER AND DAG
CC BINDING DOMAIN.
CC -I- SIMILARITY: CONTAINS 1 GAP DOMAIN.
CC -I- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -I- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X77609; GAA54700.1; -.
DR HSSP: P08799; 1MND.
DR InterPro: IPR002219; DAG_PE-bind.
DR InterPro: IPR000048; IQ.
DR InterPro: IPR001609; myosin_head.
DR InterPro: IPR000159; RA.
DR InterPro: IPR000198; RhogAP.
DR Pfam: PF00130; DAG_PE-bind; 1.
DR Pfam: PF00612; IQ; 4.
DR Pfam: PF00063; myosin_head; 2.
DR Pfam: PF00788; RA; 1.
DR Pfam: PF00620; RhogAP; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR PRODOM: PD000355; myosin_head; 2.
DR SMART: SM00109; C1; 1.
DR SMART: SM00015; IQ; 3.
DR SMART: SM00242; MSC; 1.
DR SMART: SM00314; RA; 1.
DR SMART: SM00324; RhogAP; 1.
DR PROSITE: PS00479; DAG_PE_BIND_DOM_1; 1.
DR PROSITE: PS50081; DAG_PE_BIND_DOM_2; 1.
DR PROSITE: PS50096; IQ; 3.
KW Myosin. Repeat. ATP-binding. Calmodulin-binding. Actin-binding;
KW Cytoskeleton; Coiled coil; GTPase activation; Phospho-ester binding;
KW Zinc.
FT DOMAIN 941 1045 HEAD OR MOTOR DOMAIN.
FT DOMAIN 1046 1980 TRAIL.
FT DOMAIN 1841 1861 COILED COIL (POTENTIAL).
FT DOMAIN 1918 1948 COILED COIL (POTENTIAL).
FT DOMAIN 845 856 ACTIN-BINDING.
FT DOMAIN 958 978 IQ 1.
FT DOMAIN 981 1001 IQ 2.
FT DOMAIN 1002 1024 IQ 3.
FT DOMAIN 1025 1054 IQ 4.
FT DOMAIN 1533 1641 PHOSPHO-ESTER AND DAG BINDING.
FT DOMAIN 1673 1822 GAP DOMAIN.
FT NP_BIND 239 246 ATP (POTENTIAL).
SO SEQUENCE 1980 AA; 225035 MW; D79FEC4D0FAE0C05 CRC64;
```

```

Query Match 4.3%; Score 152; DB 1; Length 1980;
Best Local Similarity 20.0%; Pred. No. 1.3;
Matches 136; Conservativity 107; Mismatches 242; Indels 196; Gaps 33;
Qy 6 MRHS--KGHPQGGIOYKMKRNPISLKLTDNRPEKSKCKPLMGKVPYLDLPSTVI--- 60
```

```

Db 1030 IRLSLCRLHQ-----RRSFQOMLEKOKAQOANETVAGAMSGE-----PSPVAGE 1078
OY 61 --SEKIQDIDIGRVEEFLSKDISYLSNKKKEAFQOTLGRISPPSPESAYTAETS 118
Cc 1079 QPSEHPVDEPDELQVETETMMNSKSPNGISPKKE-----PSPSEMTPAOKTV 1126
OY 119 PPSHDGSGFPSTVQVLSRGKLLVEKAIKHDHFLPSNLSLNSMGVYKLIHIDIRY 178
Cc 1127 PAESHE-----KVPS-----SREK-----RRSRQRGLE-----H 1151
OY 179 IEOKKELLYLKKSSTVRGCGKRVGSGAQKTRORLKKPPVK-VEDMSOLYRPFYLOLT 237
Cc 1152 VEROKKHIGOSRENSLGRPRSKASLETGESFPDPTKEPREHDLFTWTETAPSCPK-- 1209
OY 238 NMPTINYSIQKPCSPVYDKSPSSMOKOTQY-----KLRIQTGDKYGGTSTQLOLAKK 290
Cc 1210 QVPIVG---DPPKSPPLQRPASLDLSRVSPVLPSSLSLSPQDEDK-GENSTKVDKPE 1265
OY 291 KKKGYCECCLOKYEDELTHLSE-----OHRNFAQSNQYQVDDIVSKLVDFVEYKQD- 344
Cc 1266 SPGSGTQ--IQRYOHPDTERLATVAEIMWGKKLASAMLSQSLD-----LSEKPR 1312
OY 345 -----TP-KKKRIKYS--VGSLSFVSASVLAKTQEKKEVLOHISQND----- 384
Cc 1313 TAGAALPTTEERRRISFSTSVSKLSPVKTSTVEDGLSAKKRPAKHKKKSSEDPSAGPDAGL 1372
OY 385 ---COEDPTVKEQNFLYKFTQTEKKLFISEPIPHPSNELRGLNKKMSKMSLSTAE 441
Cc 1373 PTGSGQDGSKSAFKRLFLHK---AKDK-----PS-LGCVGE-----TE 1406
OY 442 DDIRONFQULPHKKNKOECILDI-----SEHTLSENDLEELRVHYK---CNIQASVHVS 493
Cc 1407 GSGGQAQGEAPARKT-----LDVPSQQRHHTTGKPLKGGKNNRNVGQITVSEKWRSS 1461
OY 494 DPTSDNGSGQKOKSDTVLPAKDKLEKDLHSIFTHDSGLITINSOEHLYQAAPRHT 553
Cc 1462 VFRKITTANLEKFLDEFLKKNVNDLSQ-----KTPLESFILATPRFAS 1506
OY 554 PPEENEDCFRNMDSLPKGIHKVYIILGRNKENLE---PNAEFKREFFTQENMI 610
Cc 1507 -----NITMTISVNGKIHVGYNDLM-----ENVQIVVSNLAER-----GENDTML 1548
OY 611 CSSPVQSLDLFQTSSEKSEPE 631
Cc 1549 VLNWFQSLDDEFTRSYNKTFD 1569
Db 1549 VLNWFQSLDDEFTRSYNKTFD 1569

RESULT 12
YAF3_SCHPO STANDARD; PRT: 1044 AA.
AC 009857; QPPTV0;
DT 01-FEB-1996 (Rel. 33, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DN HYPOTHETICAL 119.1 KDA PROTEIN C29E6.03C IN CHROMOSOME 1.
GN SPAC29E6.03C OR SPAC30.07C.
OC Schizosaccharomyces pombe (Pisition yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC NCB1_TaxID=4896;
OY 111
Cc 111
Rp SEQUENCE FROM N.A.
Cc STRAIN=972;
Ra Jones L., Murphy L., McNeil A., Simpson I., Harris D., Barrell B.G.,
Ra Rajandream M.A., Walsh S.V.,
Ra Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
Rp [2]
Rp SEQUENCE FROM N.A.
Rp STRAIN=972;
Ra McDougall R.C., Rajandream M.A., Barrell B.G., Brown S., Murphy L.,
Ra Jones L., McNeil A., Harris D.;

```

```

RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
Cc -I SIMILARITY: BELONGS TO THE VDP/USO1/YBL047C FAMILY.
Cc CC
Cc CC This SWISS-PROT entry is copyright. It is produced through a collaboration
Cc between the Swiss Institute of Bioinformatics and the EMBL outstation -
Cc the European Bioinformatics Institute. There are no restrictions on its
Cc use by non-profit institutions as long as its content is in no way
Cc modified and this statement is not removed. Usage by and for commercial
Cc entities requires a license agreement (See http://www.isb-sib.ch/announce/
Cc or send an email to license@sib-sib.ch).
Cc CC
DR EMBL: Z66525; CA91425.1; ALT_SEQ.
DR EMBL: AL136538; CAB66466.1;
KW Hypothetical protein; Coiled coil.
FT DOMAIN 600 1014 Coiled coil (POTENTIAL).
SQ SEQUENCE 1044 AA; 119127 MW; 52ACEDE174725A59 CRC64;

```

Query Match 4.38; Score 151.5; DB 1; Length 1044;  
 Best Local Similarity 19.58; Pred. No. 0.6; Indels 201; Gaps 34;  
 Matches 130; Conservative 122; Mismatches 214;

```

OY 52 YLDPSTYISEKLOKIDIGRVEEFLSKDISYLSNKKKEAFQOTLGRISPPSPESA 111
Cc 418 YLDE--VHLSDDQDYD-----NMFTSTLYTLVIDNQRRYL-----LCSIFLQDM 462
OY 112 YTAETSPHPSHDGSGSPDYVCLSRGKLV-----EKAIKD----- 149
Cc 463 DNDEDESE---SEDKVTF-----IQCVST-KLIATLRHENALQNCVGYTLILALVYGNPDS 514
OY 150 -HDFPNSIL-----SNALSMGVKLIHIDIRY--IEO--KKKELYLKK 191
Cc 515 VKDLSESSILQFLFALMDESSANSYIOGMIAVFLSLVYYICPIESPVSADY---- 570
OY 192 STSVRGCGKRVGSGAQKTRGRKKPVEKMSOLYRPFYLOLNMFFINYSIQKPCS 251
Cc 571 -----NATSAVKRD-VFINRQRLRR--NMLVEIRFL----- 600
OY 252 PFVDKSSMOKOTQYK-RIQTDGDKYGGTSTQLOLKE-----KKKGYCEC 298
Cc 601 -----SMOKQOLKSLREIDNTR--EALDSYKENSIOBEKLNLSLKTSTNLE 649
OY 299 CLQKYEDLETHLSEQR-----NFAQSN-----QYQVDDIVSKLVDFV-ETEK 343
Cc 650 QTOLEKXYHBEILDNQOKLYDIRIELDYTKNSCKQMEQVLRGEHSEIKDFTIEBSK 709
OY 344 DTPKKRIKYSVGLSPVSASVLAKTQEKQEV--ELOHISQKDCQ-EDDTVKEQNFLYK 400
Cc 710 LTKQLDIDIKNQFGIISKNRNDLSLEIKSKSLNNSLALESKKNKLENDLILTER-LNK 768
OY 401 ETOETEKLLFISEPIPHPSNELRGLNKKMSKMSLSTAEEDIRONFT-----QLPL 453
Cc 769 KNATDS-----FKNTIREAELESKKALNDLNGKNEITSDLNKLTSESTRLOLSQNLQ 824
OY 454 HKNOECILD-----ISEHTLSENDLEELRVHYDKC-NIQASVHSDSTNSQSO 503
Cc 825 DKNOLETNLRIAAADELSSMESINKQANLEKLAOKCSNLOEKI-----NFGNK 876
OY 504 -PKQSDTVLFPKQDK-----EKDHSIFTHDSGLITINSOEHLYQAAPRHT 553
Cc 877 LAKETHTKISLKLDEAATKATASLSKELTKVSEKDSLSKVSNDQK----- 925
OY 554 PPEEPNECDTRKND-----SLPSGKIHR-KVITILGRNKENLEPNAEFKREF 602
Cc 926 KEKSVNNEKFEVSOALAEANLEKLNARDEITERLKYDIGLQNASINMOSLSDSNKRTIS 985
OY 603 ITQENR 609
Cc 986 DLESKNK 992
Db 986 DLESKNK 992

```

RESULT 13  
 MY3\_SCHPO



KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF  
 MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on  
 CC use by non-profit institutions as long as its content is not in  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

-----  
 CC EMBL: X03831; CAAT7446.1; .  
 CC PIR: A23386; SAZOGM.  
 CC InterPro: IPR000561; EGF-like.  
 CC Malaria: Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;  
 CC Transmembrane; GPI-anchor.  
 CC KW SIGNAL 1 19 POTENTIAL SURFACE PROTEIN 1  
 CC CHAIN 20 1726 MEROZOITE SURFACE PROTEIN 1  
 CC FT CAROBYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 CC FT CAROBYD 272 272 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 CC FT CAROBYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 CC FT CAROBYD 567 567 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 CC FT CAROBYD 638 638 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 CC FT CAROBYD 827 827 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 CC FT CAROBYD 924 924 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 CC FT CAROBYD 944 944 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 CC FT CAROBYD 990 990 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 CC FT CAROBYD 1016 1016 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 CC FT CAROBYD 1016 1016 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 CC FT CAROBYD 1221 1221 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 CC FT CAROBYD 1613 1613 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 CC FT SEQUENCE 1726 AA; 196197 MW; DD8AD45FA352ACF3 CMC64;

Query Match 4.3%; Score 149.5; DB 1; Length 1726;  
 Best Local 20.4%; Pred. No. 1.4; Mismatches 224; Indels 223; Gaps 37;  
 Matches 141; Conservative 104;

65 QNDIDGGRVE-----FLSKDISYLSNKKKAKFNOTGRTSPVPSPESAYVFTTS 118  
 DB 309 QYDLSYKQLEAHNLISYLEKRIOTL--KKENIKELLDKIN-----EIKN 354  
 QY 119 PPHSDGSFSPSPOTVLSRGKLLV--EKAIRDHFIPNSILSNLSMGVILHID--D 174  
 DB 355 PPHASGN--TPNTL-LDKNNKIEHEEKIE-----IAKTIKFNIDSLFTDPLE 401  
 QY 175 IRYIEOKKKELYLLKSSVSVDGKRVGSAQKTRGLKPFVVEDMSQLYPTL 234  
 DB 402 LKYLKRNKKKVDYTPKSODPTS-----V 426  
 QY 235 QLTNMPFINYSIKRPSPFVY-----DKPS--SMOKOTOYKL--RIOTDGDK--- 277  
 DB 427 QIKKVIYFN-GIVYPLPLTDLHNSLADNDKNSYGDLMNPOTKKEIKIITNKKRIE 485  
 QY 278 YGTSYQQLQLEKKK--KKGCECCLOKYE---DLETHLSSEGRNPAOSNGYOVADI 330  
 DB 486 INNKKQIDLEKRIINHTKEDNKKLEDEYKSKDYBELLEFTEFMKFNPNPKDADVKI 545  
 QY 331 VSKLYVDEFEYKEDPKKK-RIKYSVGSLSPVASATKTBKKEVELHISQKDCQEDD 389  
 DB 546 FS-----ARTYNEKORYNNKSSNSVYVQKLKA-----LSYLEDYSLRKG- 591  
 QY 390 TYVKEONFLYKETOET--EKKLLEISPIPHSPENELRGLENKMSKMSKMLSTADDIRON 447  
 DB 592 --LSEDFNHTTTLTKGLLEADIKITBELIKSSNKI-----LEKN 629  
 QY 448 FTQPLAKKKQCIIDISEHTLSF-----NDLEFLRDYHKKNIOASVHVSFSDNS 500  
 DB 630 FKGLTISANASLEVIDYKLVQVALLKRIEKLIELFLKNALKMSHIVPIYRQK 669  
 QY 501 GSDP-----KKSPTV--LFPKDLKEDD--LSTI-----FTHSGILITNS 539

DB 690 KPEPYYLLVKKKDYKLEKFIKPKYKMDLKKKQAVLSITOPLVASSTEDGSHYTL 749  
 QY 540 Q-----EHLVQAKAPFT-----PPEE-----PNECDFKMDSLPGKIHRRVKII 581  
 DB 750 QSEGEVVEEPEPEPEVGHHTTYITLPPKRYKVENSIKKNSDN----- 796  
 QY 582 LGRNRENEPNAEEDKREPTIOENRIC-----SPVQSLDLFO-TSEKSEPL 632  
 DB 797 -----SQALTIVYLYKRIDEFELT--KSYICHKXIIYVSSMDOKILEVYLTPEENEEL- 848  
 QY 633 GFTSYEKSGICNVLDIWEENSNDLTAFFS 664  
 DB 849 -----KS--CDPLDLL--FNIONNIPAMYS 869

RESULT 15  
 MSPL\_PLAEP STANDARD: PRT: 1726 AA.  
 ID MSPL\_PLAEP  
 AC P50495;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)  
 GN (PMSA) (GPI95).  
 OS Plasmodium falciparum (isolate Palo Alto / Uganda).  
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 NC1\_Taxid=57270;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89005525; PubMed=3049134;  
 RA Chang S.P., Kramer K.J., Yamaga K.M., Kato A., Case S.E.,  
 RA Siddiqui W.A.;  
 RA "Plasmodium falciparum: gene structure and hydropathy profile of the  
 RA major merozoite surface antigen (gpi95) of the Uganda-Palo Alto  
 RA isolate."  
 RL Exp. Parasitol. 67:1-11(1988).  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR  
 CC (POTENTIAL).  
 CC -1- PPM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42  
 CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF  
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on  
 CC use by non-profit institutions as long as its content is in no  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

-----  
 CC EMBL: M37213; AAA29611.1; .  
 CC InterPro: IPR000561; EGF-like.  
 CC Malaria: Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;  
 CC Transmembrane; GPI-anchor.  
 CC KW SIGNAL 1 19 POTENTIAL SURFACE PROTEIN 1  
 CC CHAIN 20 1726 MEROZOITE SURFACE PROTEIN 1  
 CC FT CAROBYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 CC FT CAROBYD 272 272 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 CC FT CAROBYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 CC FT CAROBYD 567 567 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 CC FT CAROBYD 638 638 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 CC FT CAROBYD 827 827 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 CC FT CAROBYD 924 924 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 CC FT CAROBYD 944 944 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 CC FT CAROBYD 990 990 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 CC FT CAROBYD 1016 1016 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 CC FT CAROBYD 1114 1114 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 CC FT CAROBYD 1221 1221 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 CC FT CAROBYD 1613 1613 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 CC FT SEQUENCE 1726 AA; 196174 MW; 5B59CEBFA2FA026 CRC64;



**THIS PAGE BLANK (USPTO)**